



# Digital approaches to paleography and book history: some challenges, present and future

Peter A. Stokes\*

King's College London, London, UK

**Keywords:** paleography, book history, diplomatic, digital humanities, digital methods, typography, codicology

## OPEN ACCESS

### Edited by:

Lior Wolf,  
Tel-Aviv University, Israel

### Reviewed by:

Daniel Stökl Ben Ezra,  
Ecole pratique des hautes études  
(EPHE-Sorbonne), France

### \*Correspondence:

Peter A. Stokes  
[peter.stokes@kcl.ac.uk](mailto:peter.stokes@kcl.ac.uk)

### Specialty section:

This article was submitted to Digital Paleography and Book History, a section of the journal Frontiers in Digital Humanities

**Received:** 08 June 2015

**Accepted:** 08 October 2015

**Published:** 29 October 2015

### Citation:

Stokes PA (2015) Digital approaches to paleography and book history: some challenges, present and future. *Front. Digit. Humanit.* 2:5. doi: 10.3389/fdigh.2015.00005

Quantitative methods and the use of technology in paleography and book history are by no means new. The New Palaeographical Society took advantage of the latest printing techniques to produce its albums of facsimiles in the late nineteenth century (Thompson et al., 1903–1930); Jean Mallon began using film in the 1930s to illustrate the development of script (Pouille, 1977); Gilissen (1973) was applying statistical measures to script long before widespread access to computers; and even Mabillon (1681) used advanced printing methods in his foundational study of 1681. Smith (1938) applied innovations in photography for the recovery of damaged or illegible script in the 1930s, and Malachi Beit-Arié's database of Hebrew manuscripts began in the 1960s and remains one of – probably – the most important works of quantitative codicology. This list is by no means complete, and it demonstrates a continued interest in finding new ways to help us understand books and documents. However, it is also clear that recent developments have dramatically transformed the field. It is now 10 years since publication of what is perhaps the seminal article in what has since become known as “digital” or “computational” paleography (Ciula, 2005). Much excellent work has been done since then, and the whole field of paleography has perhaps been reinvigorated as a result [see, for instance, Rehbein et al. (2009), Fischer et al. (2011), Nelson and Terras (2012), Hassner et al. (2013)]. Nevertheless, much remains to be done, both within paleography in the narrow sense and also, probably even more so, in other areas of book history and diplomatic.

Regarding paleography, much work in the last 10 years has been done on questions of “who,” “when,” “where,” and “what”: what was written, by whom, when, and where? However, there is very much more to paleography than this. What about questions of technique, such as searching for “stabbing” strokes that might indicate a scribe used to writing on wax, or changes in scribal “equilibrium,” which might suggest expertise, forgery, or imitation (Stokes, 2011, 2014)?

## “GLOBAL” APPROACHES

Most work, to date, has focused on a few “major” writing systems. One challenge is therefore to ensure that the full diversity of writing and book production is represented in this work, and that we do not perpetuate existing divides of culture, politics, history, and indeed power. Beyond this, interest is also growing in what might be called “transversal” or even “global” paleography<sup>1</sup>. Paleography and Book History have both long been multi- and interdisciplinary; nevertheless, scholars tend to be very specialized in what that they consider. This is for good reason, given the advanced training one needs to understand the writing systems, languages, techniques of book production, historical and cultural context of production, and so on. However, some digital methods

<sup>1</sup>The term “transversal palaeography” is taken from the Groupe de recherches transversales en paléographie at the École pratique des hautes études in Paris. Comparable also is the Center for the Study of Manuscript Cultures in the University of Hamburg, and Derolez (2003a).

are already proving effective across very different writing systems, and comparative codicology has existed for some time [Hoffman and Hunzinger (1998); see also Bausi et al. (2015)], so how far can these new methods be applied “transversally”? What are the opportunities and limits? Can methods for, say, modern printed Hebrew apply also to Mongolian Vertical script? Is Western European codicology transferrable to and from, say, Malaysian palm-leaf books? Can there be a universal model for the representation of books and handwriting? For handwriting and print? For decoration? Are modern and medieval books as different as scholarly practice seems to suggest, or can the same methods be applied to both? Can paleographers learn from typographers, and vice versa? Or is even asking these questions part of a vain “quest for universalism” (Reider and Röhle, 2012, pp. 78–9). This may prove straightforward for problems, such as writer identification, which are relatively agnostic regarding the writing system, although even these may include assumptions about the mechanics of writing which are by no means universal, such as the form of pen or brush, and the way it is held and manipulated (Schomaker, 2008). However, the challenge here seems much greater for more semantic or ontological approaches [as suggested, for instance, by Hassner et al. (2013)]. Finding universal elements here may prove impossible, but success just might hint at something very fundamental about humanity and our development and use of technology.

## BEYOND PALEOGRAPHY

As already noted, some of the earliest uses of computers for manuscript studies was in the field of codicology, and work in that area has continued ever since. However, the focus of the last 10 years seems to have been paleographical. There are many notable exceptions, such as continued work in quantitative codicology, the series of Digital Diplomatics conferences, the application of IconClass to online frameworks for the study of Art History, the application of RTI to inscriptions, and the representation of books, texts, and other related entities through ontologies, such as FRBRoo, CIDOC-CRM, TEI, and EAD. Similarly, the study of printed books has also benefited significantly from digital approaches through OCR, layout analysis, and others. Nevertheless, there seem to be some clear omissions. How can we fully integrate the text with a model of the construction of the book that it is preserved in, incorporating not only the page layout but also the binding structure, pages that are (or were) physically part of the same sheet of paper or parchment, and so on? Books and sometimes documents are not static objects but often change in content, binding, and even structure, so how can this be captured and represented? If “the medium is the message,” as is so often asserted, then how can the physical medium of the book or document be represented in a digital edition of the text? To what extent can or should study of books and documents in digital form engage with theories of new media and remediation, for instance, those of Bolter and Grusin (2000)? What about more three-dimensional forms of written communication, which are not readily amenable to photography, such as inscriptions, clay tablets, or cylinder seals? What can we learn by bringing librarians, archivists, conservationists, even publishers, and others

into discussion with paleographers, codicologists, diplomatists, and book historians?

## THE “BIG QUESTIONS” OF DIGITAL HUMANITIES

Many of the “big questions” of Digital Humanities are very relevant to book history and diplomatic and have been discussed already in this context, particularly by Hassner et al. (2013), and this paragraph and the next draw heavily on that discussion. A big and perhaps unanswerable question is if we can find the limits of automation, now and in the foreseeable future. Another is the “80/20” question: given the inherently messy and “human” nature of book production and diplomatic, it seems unlikely that digital methods will produce perfect results for many of our questions at any time soon. As well as attempting to improve the accuracy of the results, then, it is also worth asking what we can usefully do now with results that are “near enough.” One example is work on finding joins in fragments from the Cairo Genizah: this corpus of many thousands of fragments is too big for people to manage alone, but the computer can make suggestions of where to look (Wolf et al., 2011a,b; Genizah, 2015). Even if only 1% of these suggestions are accurate, this still equates to some 3,500 new joins, which had not been found before. In contrast to this is writer verification, where the methods might be the same but the underlying question is importantly different: not “find me specimens which are similar,” but rather “I have two specimens of writing, now tell me if they were written by the same person.” The difference here is how to verify the result. These two samples are somehow similar, which is straightforward enough to determine given metrics for similarity. However, it is not at all clear how to verify that the similarity identified by the computer necessarily means production by the same writer, particularly for cases in which we have no substantial ground truth against which to test our methods. Assuming that we do not have a perfectly reliable algorithm, any result will be provided in the form of a likelihood, and it will then be incumbent on the human expert to make the final decision. However, if the experts were able to verify this themselves then they would not have needed the computer. On what basis can this decision be made? How can we “cross-examine” the computer, a question raised by Davis some years ago and still only barely addressed [Davis (2007), p. 266 no. 27; see also Schomaker (2008), p. 248]? Is favoring the computational approach simply succumbing to the “lure of objectivity” (Reider and Röhle, 2012, pp. 71–3)?

This discussion leads into questions about evidence and communication. How can we address the “black box” problem (Hassner et al., 2013, pp. 25–6), and ensure that results can be verified? How do we present computational results in ways that the audience can understand and use them? Does this matter: to what extent must all specialist scholarship be accepted at least partly on trust? Strikingly relevant here is the question that Derolez raised more than 10 years ago: “how is it possible to proceed in such a way that the description of handwriting is as clear and convincing to its reader as it is to its author?” (Derolez, 2003b, p. 7). His question was not “how can I know,” but rather “how can I convince you,” and this distinction seems often to be forgotten.

These questions in turn suggest the importance of visualization and interface design, areas that seem to have received relatively little attention when applied to books and manuscripts. Is the seemingly ubiquitous manuscript viewer really the most appropriate way of presenting this material? Are images really the best way to “reproduce” the original at all (cf. Tarte, 2011, p. 7, and Reider and Röhle, 2012, pp. 73–5)? To what extent can or should other approaches, such as structured data, also be considered representations in this way? How can we communicate the non-visual aspects of our materials? How can search queries and results be improved in a field with widely varying terminology, with wide uncertainty in date range and place, and where the objects themselves have changed significantly over time?

Finally, this diachronic aspect is another challenge that seems to have received relatively little attention. Stansbury has already noted that digital methods tend to focus on classification rather than development or evolution (Stansbury, 2009, pp. 237 and 247–8), but his challenge has met with relatively little attention. Some work has been done on dating manuscripts based on the assumption that variations in script develop over time (He et al.,

2014), but the objective here is still to determine dates rather than to investigate the evolution *per se* (but cf. Poulle, 1977). How can we use computers to help represent and investigate diachronic variation as an end in itself? How can we even refer unambiguously to a single manuscript when that manuscript may be composite, may have been rebound, had pages rearranged, sections moved between different volumes, and so on? Places change as well: the London of today is in some respects an entirely different entity from that of the eleventh century, for example. What are the implications of all this for linked data and the semantic web? Is your London really the same as mine? And what are the consequences if it is not?

The questions raised here are not new, and indeed many of them have been considered before, whether in Digital Humanities or long before that label was even imagined. Nevertheless, the challenges remain. The stakes are high, insofar as effective responses to these questions would have very wide implications not just for the study of books and documents but much more widely. A lot has happened since Ciula’s seminal article, but a lot remains to be done.

## REFERENCES

- Bausi, A., Borbone, P.G., Briquel-Chatonnet, F., Buzi, P., Gippert, J., Macé, C. et al. (2015). *Comparative Oriental Manuscript Studies: An Introduction*. Hamburg: Comparative Oriental Manuscript Studies. Available at: <http://www1.uni-hamburg.de/COMST/handbookonline.html>
- Bolter, J.D., and Grusin, R. (2000). *Remediation: Understanding New Media*. Cambridge, MA: MIT Press.
- Ciula, A. (2005). Digital palaeography: using the digital representation of medieval script to support palaeographic analysis. *Digit. Medievalist* 1: 1.
- Davis, T. (2007). The practice of handwriting identification. *The Library, 7th series* 8: 251–76. doi:10.1093/library/8.3.251
- Derolez, A. (2003a). *The Publications Sponsored by the Comité International de Paléographie Latine*. Available at: <http://www.palaeographia.org/cipl/derolez.htm>
- Derolez, A. (2003b). *The Palaeography of Gothic Manuscript Books: From the Twelfth to the Early Sixteenth Century*. Cambridge: Cambridge University Press.
- Fischer, F., Fritze, C., and Vogeler, G. (eds) (2011). *Kodikologie und Paläographie im Digitalen Zeitalter 2 – Codicology and Palaeography in the Digital Age 2*. Norderstedt: Books on Demand.
- Genizah. (2015). Conferences and papers. In *The Freidberg Genizah Project*. Available at: <http://www.genizah.org/ProfessionalPapers.aspx>
- Gillissen, L. (1973). *L'expertise des écritures médiévales. Recherche d'une méthode avec application à un manuscrit du XIe siècle: le Lectionnaire de Lobbes, codex Bruxellensis 18018*. Ghent: Éditions scientifiques E. Story-Scientia.
- Hassner, T., Rehbein, M., Stokes, P.A., and Wolf, L. (2013). Computation and palaeography: potentials and limits. *Dagstuhl Manifestos* 2: 14–35. doi:10.4230/DagMan.2.1.14
- He, S., Samara, P., Burgers, J., and Schomaker J. (2014). *Towards Style-Based Dating of Historical Documents*. Paper Presented at the International Conference on Handwriting Recognition [ICFHR-2014]. Crete. Available at: [http://www.researchgate.net/publication/263518853\\_Towards\\_style-based\\_dating\\_of\\_historical\\_documents](http://www.researchgate.net/publication/263518853_Towards_style-based_dating_of_historical_documents)
- Hoffman, P., and Hunzinger, C. eds. (1998). *Recherches de codicologie comparée: la composition du codex au Moyen Age, en Orient et en Occident*. Paris: Presses de l'Ecole normale supérieure.
- Mabillon, J. (1681). *De re diplomatica*. Paris: sumtibus Ludovici Billaine.
- Nelson, B., and Terras, M. eds. (2012). *Digitizing Medieval and Early Modern Material Culture*. Tempe, AZ: Center for Medieval and Renaissance Studies.
- Poulle, E. (1977). Une histoire de l'écriture. *Bibliothèque de l'École des Chartes* 135: 137–44. doi:10.3406/bec.1977.450098
- Rehbein, M., Sahle, P., and Schaßan, T. (eds) (2009). *Kodikologie und Paläographie im Digitalen Zeitalter – Codicology and Palaeography in the Digital Age*. Norderstedt: Books on Demand.
- Reider, B., and Röhle, T. (2012). Digital methods: five challenges. In *Understanding Digital Humanities*, ed. D.M. Berry, 67–84. Houndmills: Palgrave Macmillan.
- Schomaker, L. (2008). Writer identification and verification. In *Advances in Biometrics: Sensors, Algorithms and Systems*, ed. N.K. Ratha and V. Govindaraju, 247–264. London: Springer.
- Smith, A.H. (1938). The photography of manuscripts. *London Mediev. Stud.* 1: 179–207.
- Stansbury, M. (2009). The computer and the classification of script. In *Kodikologie und Paläographie im Digitalen Zeitalter – Codicology and Palaeography in the Digital Age*, ed. M. Rehbein, et al., 237–249. Norderstedt: Books on Demand.
- Stokes, P.A. (2011). The problem of grade in post-Conquest Vernacular minuscule. *New Mediev. Lit.* 13: 23–47. doi:10.1484/J.NML.1.102438
- Stokes, P.A. (2014). Describing handwriting – again, in Digital palaeography: new machines and old texts. *Dagstuhl Reports* 4: 127–128. doi:10.4230/DagRep.4.7.112
- Tarte, S. (2011). *Digital Visual Representations in Papyrology: Implications on the Nature of Digital Artefacts*. Available at: [https://www.academia.edu/4886465/Interpreting\\_Textual\\_Artefacts\\_Cognitive\\_Insights\\_into\\_Expert\\_Practices](https://www.academia.edu/4886465/Interpreting_Textual_Artefacts_Cognitive_Insights_into_Expert_Practices)
- Thompson, E.M., Warner, G.F., Kenyon, E.G., and Gilson, J. P. (1903–1930). *The New Palaeographical Society Facsimiles of Ancient Manuscripts. First and Second Series*. Oxford: Oxford University Press.
- Wolf, L., Littman, R., Mayer, N., German, T., Dershowitz, N., Shweka, R., et al. (2011a). Identifying join candidates in the Cairo Genizah. *Int. J. Comput. Vis.* 94: 118–35. doi:10.1007/s11263-010-0389-8
- Wolf, L., Dershowitz, N., Potikha, L., German, T., Shweka, R., and Choueika, Y. (2011b). Automatic paleographic exploration of genizah manuscripts. In *Kodikologie und Paläographie im Digitalen Zeitalter 2 – Codicology and Palaeography in the Digital Age 2*, ed. F. Fischer, et al., 157–179. Norderstedt: Books on Demand.

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Stokes. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.